Estimating grassland biomass using SVM band shaving of hyperspectral data

Clevers, J G P W ; van der Heijden, G W A M ; Verzakov, S ; Schaepman, Michael E

Abstract: In this paper, the potential of a band shaving algorithm based on support vector machines (SVM) applied to hyperspectral data for estimating biomass within grasslands is studied. Field spectrometer data and biomass measurements were collected from a homogeneously managed grassland field. The SVM band shaving technique was compared with a partial least squares (PLS) and a stepwise forward selection analysis. Using their results, a range of vegetation indices was used as predictors for grassland biomass. Results from the band shaving showed that one band in the near-infrared region from 859 to 1,006 nm and one in the red-edge region from 668 to 776 nm used in the weighted difference vegetation index (WDVI) had the best predictive power, explaining 61 percent of grassland biomass variation. Indices based on short-wave infrared bands performed worse. Results could subsequently be applied to larger spatial extents using a high-resolution airborne digital camera (for example, Vexcel's UltraCamTM).

DOI: https://doi.org/10.14358/PERS.73.10.1141

Originally published at:
Clevers, J G P W; van der Heijden, G W A M; Verzakov, S; Schaepman, Michael E (2007). Estimating grassland biomass using SVM band shaving of hyperspectral data. Photogrammetric Engineering Remote Sensing, 73(10):1141-1148.
DOI: https://doi.org/10.14358/PERS.73.10.1141

Estimating Grassland Biomass Using SVM Band Shaving of Hyperspectral Data

J.G.P.W. Clevers, G.W.A.M. van der Heijden, S. Verzakov, and M.E. Schaepman

Abstract

In this paper, the potential of a band shaving algorithm based on support vector machines (SVM) applied to hyperspectral data for estimating biomass within grasslands is studied. Field spectrometer data and biomass measurements were collected from a homogeneously managed grassland field. The SVM *band shaving technique was compared with a partial least squares (PLS) and a stepwise forward selection analysis. Using their results, a range of vegetation indices was used as predictors for grassland biomass. Results from the band shaving showed that one band in the near-infrared region from 859 to 1,006 nm and one in the red-edge region from 668 to 776 nm used in the weighted difference vegetation index (WDVI) had the best predictive power, explaining 61 percent of grassland biomass variation. Indices based on short-wave infrared bands performed worse. Results could subsequently be applied to larger spatial extents using a high-resolution airborne digital camera (for example, Vexcel's UltraCam™).*

Introduction

For efficient grassland management, information on the grassland status and its spatial variation within fields is of the utmost importance. Farmers need such information at early growth stages for guiding the fertilizer supply within fields in order to achieve optimal growth. Towards the end of the growing phase, such information is required for an early yield prediction and determination of the optimal time of harvest. Currently management is mainly based on qualitative expert knowledge. Quantitative information on the actual status of grass swards at the right moment in the season is preferred. Traditionally, destructive field measurements have been very tedious, time-consuming and not applicable to large areas. However, they have been used extensively to calibrate local, statistical-based models. In order to map the spatial variability of vegetation variables over large areas, airborne or spaceborne instruments are more suitable. A major constraint for applying remote sensing techniques is that the variability of the biophysical and chemical variables of grass swards on farmer's fields in practice may be quite limited. Much remote sensing research has been performed on experimental fields with a large variability in biophysics of the canopy. In this study, we will focus on spectral information that can be used to predict spatial patterns even when the grass inherent variation of structure and biochemistry is limited.

Both statistical and physical approaches have been used for describing the relationship between spectral measurements and biophysical variables of vegetation. Statistical methods are based on regression models using either the original spectral bands or transformations on the spectral bands as independent variables. For the latter, a whole range of vegetation indices has been developed for estimating variables like biomass and leaf area index for a range of vegetation types (Broge and Leblanc, 2001; Daughtry *et al.*, 2000; Haboudane *et al.*, 2004; Haboudane *et al.*, 2002; Schlerf *et al.*, 2005; Thenkabail *et al.*, 2002). Traditionally, most vegetation indices have been broadband indices based on red and near-infrared (NIR) bands. For estimating leaf chlorophyll and nitrogen concentration narrow spectral bands (of 10 nm width or less) have shown to be important, for instance for deriving information on the red-edge position (Broge and Leblanc, 2001; Clevers and Jongschaap, 2001). Physical-based methods often use radiative transfer models describing the interaction of radiation with the plant canopy based on physical principles. Subsequently, model inversion is used for estimating biophysical properties of the canopy (Atzberger, 2004; Combal *et al.*, 2003; Jacquemoud *et al.*, 2000).

One problem that arises when using hyperspectral data as explanatory variables is the high dimensionality: they require many observations to achieve a good description of the distribution in high dimensional space (the so-called *curse of dimensionality*). Thus, the variance-covariance matrix is generally poorly estimated, and this can lead to worse prediction results compared to using less explanatory variables. Besides, hyperspectral data often show a high degree of collinearity of neighboring bands. This collinearity decreases the intrinsic dimensionality of the dataset, which would mean that fewer observations are required. However, nearly linearly-related variables make the variance-covariance matrix nearly singular. This leads to problems with matrix inversion and yields highly unstable parameter estimates. Generally, feature reduction (e.g., principal components transformation and partial least squares) or feature selection (e.g., stepwise linear regression) techniques are applied to

J.G.P.W. Clevers and M.E. Schaepman are with the Centre for Geo-Information, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, The Netherlands (jan.clevers@wur.nl).

G.W.A.M. van der Heijden is with Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands.

S. Verzakov is with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

Photogrammetric Engineering & Remote Sensing
Vol. 73, No. 10, September 2007, pp. 1141–1148.

0099-1112/07/7310–1141/$3.00/0

overcome these problems (Bruzzone and Serpico, 2000). The advantage of feature selection is that the derived features are easily interpretable from a physical point of view. This is not the case for, e.g., the principal components obtained from hyperspectral data. When feature selection techniques reduce the dimensionality of hyperspectral data, they should preserve the key information content within a few spectral bands.

A recent development in the field of feature selection are methods known as shaving (Verzakov *et al.*, 2004). Shaving methods for spectroscopic data first detect correlation between spectral bands, and based on these correlations, the feature selection is performed. Final features are averages of continuous bands of wavelengths. As stated before, such features are easily interpretable from a physical point of view (Verzakov *et al.*, 2004). In this paper, we apply a modification of a shaving algorithm based on support vector machines (SVM). Subsequently, the features (possibly narrow or broad spectral bands) can be implemented in a simple regression model.

The objective of the present study is to compare different spectral-based approaches for predicting biomass of grassland. The main goal is the derivation of the optimal spectral bands based on an SVM-modified, band shaving algorithm. Results will be compared with partial least squares and stepwise regression techniques. The latter techniques do not use the fact that neighboring wavelengths are highly correlated, thus neglecting this useful *a priori* information. In addition to regression models using the optimal spectral bands, various vegetation indices, which make use of the previous results, will be tested. The various phases of the study include:

- To assess the overall potential of hyperspectral data in a predictive model using partial least squares (PLS) techniques,
- To select the optimal narrow-band spectral bands by a stepwise linear regression technique,
- To define the optimal spectral bands by a SVM band shaving technique. This may not only result in narrow-band features, but also in broad-band features, and
- To test the potential of vegetation indices as predictors for grassland biomass.

Data Description

Field Data

At the "Droevendaal" experimental farm in Wageningen (the Netherlands), a total of 20 plots were defined within a 2.2 ha grass field with a mixture of grass species and white clover. These plots were each 15 m long and 3 m wide with a spacing of approximately 10 m between the plots. Plots were harvested using a plot-harvester on 30 July 2004. The biomass was recorded with a built-in weighing unit on the harvester. After cutting, drill samples were taken from the harvested material. These samples were oven dried during 72 hours at 70°C. After drying, samples were weighed again for determination of dry matter weight.

Table 1 summarizes the destructive measurements performed for the 20 plots of the grass/clover experiment. Yield figures were at a high level for all plots. The plots were defined within a normally treated grassland field and no treatment differences existed between the various plots. As a result the range in yield figures is not widespread, which may limit the predictive power of spectral measurements. However, such a small variety is realistic, in particular when comparing with applications in precision farming (Moran *et al.*, 1997), rather than an artificially induced variability due to treatment effects (Hansen and Schjoerring, 2003).

TABLE 1. SUMMARY STATISTICS FOR THE 20 PLOTS OF THE GRASS/CLOVER EXPERIMENT (HARVESTED ON 30 JULY 2004). (RMSE = ROOT MEAN SQUARE ERROR; CV = COEFFICIENT OF VARIATION)

	Minimum	Maximum	Mean	RMSE	CV
Biomass (t/ha)	10.91	24.39	17.79	3.30	0.19
DM weight (kg/ha)	2186	4187	3346	487	0.15
DM content (%)	16.24	21.33	18.96	1.08	0.06

Field Spectrometer Data

On 29 July 2004, a field campaign with an ASD FieldSpec® Pro FR spectroradiometer (www.asdi.com) was performed. The FieldSpec® Pro spectroradiometer was deployed using a fiber optic cable with a 25° field of view. Measurement height above the plot was about 1 to 1.5 m. As a result, the field of view at the plot level was circular with a radius ranging from 0.22 to 0.33 m. About 10 measurements per plot were performed, whereby each measurement represents the average of 50 readings at the same spot. The sampling interval was 1 nm. Although weather conditions were constant with a very low atmospheric turbidity; the spectroradiometer was calibrated regularly using a Spectralon white reference panel. Noisy regions were left out from further analysis. These were the spectral regions below 450 nm, above 2,400 nm, and water absorption bands at 1,350 to 1,450 nm and 1,800 to 1,950 nm.

Methods

Partial Least Squares

Regression models with hyperspectral data as explanatory variables are often based on partial least squares (PLS) techniques (Hansen and Schjoerring, 2003; Huang *et al.*, 2004; Kooistra *et al.*, 2004). PLS techniques are capable of dealing with high dimensional input data with a high degree of collinearity. They apply regression modeling between the observed spectral data and the response variable to select a limited number of latent variables. PLS techniques are simple to use, fast, and they show a relatively good performance (Thissen *et al.*, 2004). Like with principle components, the physical interpretation of the latent variables is difficult.

In a first step, all ground spectroradiometer measurements (the explanatory variables) were used to estimate the response variables (fresh biomass and dry matter weight) using PLS models. In order to obtain an unbiased estimate of the prediction error, a nested leave-one-out (LOO) analysis was applied. The nested approach was used to select the optimal number of latent variables at the inner loop, while still having an unbiased prediction error at the outer loop. To select the number of latent variables in the PLS-model, the root-mean-square error of cross-validation (RMSECV) was calculated as:

$$\text{RMSECV} = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{n}}, \quad (1)$$

where $\hat{y}_i$ and y_i were the leave-one-out predictions and observed values of the variable of interest (the response variable), i.e., the model was fitted without using observation i, and then observation i was predicted using this model. This was performed for all observations (n) iteratively. The number of latent variables selected was chosen as parsimonious as possible, by taking the minimal number

of latent variables whose root-mean-square error of cross validation (RMSECV) was not significantly different from the overall minimum RMSECV, using a randomisation t-test ($\alpha = 0.10$) (Van der Voet, 1994).

To obtain insight in the prediction capability of the model, we use the percentage of variation accounted for by the cross validated model (using the leave-one-out predictions) with regard to the total variation in the dataset:

$$Q^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (2)$$

where $\bar{y}$ is the average value of the response variable of interest. Note that the predicted y is not used in the training and selection of the PLS model. By using an iterative approach, a prediction for every observation is obtained. The Q^2 has strong resemblance with the traditional coefficient of determination R^2 but can become negative if the prediction of the model is inadequate (e.g., in case of overfitting). For calculating R^2 $\hat{y}_i$ in Equation 2 is calculated from one regression equation using all observations instead of using the leave-one-out predictions. The Q^2 is (as R^2) strongly sensitive to the variation within the data set, so it should be considered with caution.

Stepwise Forward Selection

Reducing the dimensionality of hyperspectral imaging by selecting just a few contiguous regions may produce affordable and robust multispectral systems for airborne and spaceborne applications. A simple feature selection technique often used in hyperspectral remote sensing is the stepwise linear regression (Serrano *et al.*, 2002). With this technique the optimum set of (individual) spectral bands for estimating vegetation variables is selected. As a model becomes more complex (more variables), it is able to adapt to more complicated underlying structures in the data set, but the prediction error increases. In general, there is an optimal model complexity that gives minimum expected prediction error (Hastie *et al.*, 2001).

A stepwise forward selection approach was applied on all wavelengths of the FieldSpec® data. The number of predictors was limited to three variables, including the intercept, because of the small sample size of the experiment. The Q^2 was calculated using a leave-one-out procedure: select best set of bands on n-1 plots and use last one for the prediction error, and repeat this for all plots.

SVM Band Shaving

Then, the support vector machine band shaving algorithm as proposed by Verzakov *et al.* (2004) was used to select the best spectral bands as predictors in this experiment. The SVM band shaving algorithm is based on the SVM recursive feature elimination (SVM RFE) approach for a classification problem (Guyon *et al.*, 2002). The idea of this approach (SVM RFE) is similar to the so-called backward feature selection (BFS). BFS starts from the complete set of features and eliminates them one by one in such a way that the classification performance of the classifier, trained on the remaining features, is the best for this particular number of features. SVM RFE also starts from the complete set of features and removes them one by one. However, SVM RFE selects features to be removed in a different way. The linear SVM classifier is trained on the current feature set, which results in a set of feature weights $w_i, i = 1 \ldots p$. The feature with the smallest absolute value $|w_i|$ is considered to be the least important and is removed from the feature set. This procedure continues until the desired number of features or classification performance is reached. To speed up computations one can remove some predefined portion of features with the lowest absolute weights. In theory it is possible to use any other linear classifier. But at the first step, we have to deal with high-dimensional data, and it is desirable if not necessary to employ robust classifiers (like SVM) which do not suffer from overtraining.

In Verzakov *et al.* (2004) SVM band shaving was proposed, which is an adaptation of the SVM RFE for spectroscopic data. SVM band shaving, like SVM RFE, starts from the training of the linear SVM classifier on the complete feature set. Then, the obtained set of feature weights w_i is considered as a (discrete) function with a natural ordering of the argument i (as is the case for wavelength in spectroscopic data). The local minima of this function are expected to be boundaries between highly correlated spectral regions responsible for different sources of variation relevant for the discrimination problem. (Note, that the last statement is not a theoretical result but a heuristic definition supported by empirical studies. More discussion on this topic will be published elsewhere.) All the features situated between two neighboring minima are considered as belonging to one band and averaged out (with taking into account the sign of the corresponding w_i) to form a new feature. Of course, the problem of finding local minima for a discrete function is ill-posed, and one needs to use smoothing to compute derivatives. After extraction of the new features (bands), the standard SVM RFE is applied to this new feature set. Such a modification of the algorithm uses *a priori* knowledge about spectroscopic data: spectral values at neighboring wavelengths are highly correlated. It helps to prevent tedious work on elimination of many very similar features. And, what is more important, it automatically combines features from contiguous regions in single bands.

In this paper, we deal with regression not with a classification problem. So, the SVM band shaving had to be adapted to a regression context for this purpose. Hence, we employed linear SVM regression (Vapnik, 1998) instead of linear SVM classification.

Vegetation Indices

Vegetation indices are often used as predictors for vegetation variables in a statistical way. The exact specification of an index depends on the available spectral bands at hand. For instance, the same indices can be applied to both narrow and broad spectral bands, yielding different values for the same index on the same object. Many studies have been performed comparing the information content using narrow spectral bands from hyperspectral imagery with the information content based on traditional broad spectral bands (Schlerf *et al.*, 2005; Thenkabail *et al.*, 2002). The estimation of biophysical vegetation properties shows an improved sensitivity when using narrow spectral bands (Elvidge and Chen, 1995; Gong *et al.*, 2003; Hansen and Schjoerring, 2003).

The use of vegetation indices is the most common way to drastically reduce the dimensionality of the data set. The advantage is that also non-linear functions are used for defining vegetation indices. A cross-section of frequently used vegetation indices (VIs) was tested in this research. This included the traditional Normalized Difference Vegetation Index (NDVI) and Simple Ratio (SR) index (Rouse *et al.*, 1974), the soil-correcting indices like the Soil-Adjusted Vegetation Index (SAVI) (Huete, 1988) and the Weighted Difference Vegetation Index (WDVI) (Clevers, 1989), the atmosphere-correcting Global Environmental Monitoring Index (GEMI) (Pinty and Verstraete, 1992), red-edge (RE) indices like the Guyot method (Guyot and Baret, 1988), Lagrangian interpolation (Dawson and Curran, 1998) and the

second derivative (Demetriadesshah *et al.*, 1990), the Enhanced Vegetation Index (EVI) as used for MODIS data (Huete *et al.*, 2002), and the MERIS Terrestrial Chlorophyll Index (MTCI) as used for MERIS data (Dash and Curran, 2004). Recently, a lot of attention is also being focused on using the short-wave infrared (SWIR) region of the electromagnetic spectrum because SWIR wavelengths penetrate better into the canopy than red wavelengths. This way, saturation problems of some of the traditional vegetation indices using red and NIR wavelengths are reduced. The SWIR wavelengths are related to canopy water content that has a close correlation with canopy biomass and leaf area index (Hunt, 1991). Most relevant canopy water related indices are the Water Index (WI) (Penuelas *et al.*, 1997) and the Normalized Difference Water Index (NDWI) (Gao, 1996). Although the original NDWI used wavelengths at 860 nm and 1240 nm, the version using 1,640 nm and 2,130 nm instead of the 1,240 nm are also being used (Chen *et al.*, 2005). Indices were calculated for the 1 nm spectral bands. Subsequently, if possible, they were also calculated for the spectral bands obtained from the SVM band shaving. Again the leave-one-out procedure was applied for determining the predictive power of the vegetation indices.

Results and Discussion

The small variation of the vegetation characteristics within the grass/clover experiment studied is confirmed by the harvesting statistics in Table 1. In particular the dry matter content shows a low variability across plots. As a result we expect similar results when we try to estimate either fresh biomass or dry matter (DM) weight in this study. Figure 1 confirms that the dry matter weight and the fresh biomass are highly correlated.

Figure 2 shows the mean, minimum, and maximum of the spectral signatures for the 20 grassland plots. This figure shows that also the spectral variation between the plots is rather small. Variation is most pronounced in the NIR part of the electromagnetic spectrum. However, the NIR region also exhibits highest reflectance values. When looking at the coefficient of variation (CV), by dividing the root-mean-square error per wavelength by the corresponding mean spectral reflectance, we see that this CV is in the range of 0.04 to about 0.08 for the whole spectrum (Figure 3), when omitting the well-known atmospheric water absorption features. This figure also shows that the second water absorption feature should be extended to about 1,960 nm in order to avoid noisy measurements.

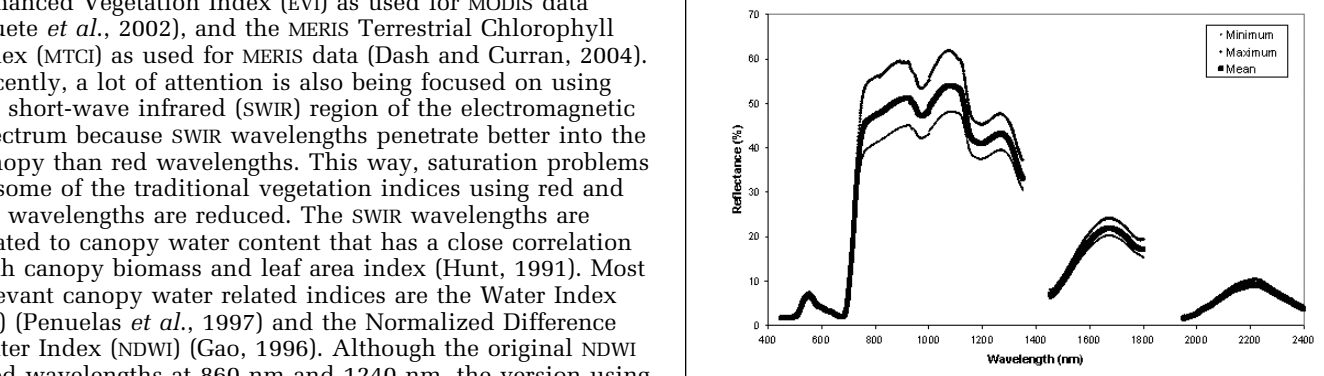

Figure 2. Mean, minimum, and maximum of the spectral reflectance for the 20 grass-clover plots.

The relationship between canopy reflectance measurements as measured with the FieldSpec® spectroradiometer and biophysical (biomass) vegetation parameters are studied by plotting the correlation coefficients (correlograms) (Figure 4). The patterns of correlation were similar for fresh biomass and dry matter weight across all wavelengths. Spectral bands in the visible part of the spectrum up to about 715 nm exhibited a negative correlation with fresh biomass and dry matter weight. In the NIR and SWIR regions, spectral bands showed a positive correlation with these characteristics. In general, the best correlations were obtained between spectral measurements and fresh biomass. The best correlation coefficient was found at 937 nm for fresh biomass and dry matter yield. In addition to the 937 nm spectral band, the 1,135 nm spectral band also showed a relatively high correlation with fresh biomass and dry matter yield (Figure 4). Both 937 nm and 1,135 nm are situated just at the lower wavelength shoulder of water absorption features situated at about 970 nm and 1,200 nm (Curran, 1989). Both wavelengths are highly correlated ($R^2 = 0.96$) for this data set. Because of the strong correlation between fresh biomass and dry matter weight, we will mainly focus on the prediction of fresh biomass further on in this study.

Partial Least Squares

Results of the prediction of the fresh biomass using all narrow-band spectral measurements obtained with the

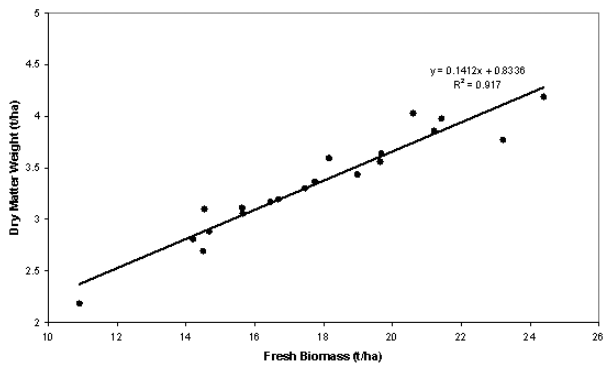

Figure 1. Relationship between fresh biomass and dry matter weight for the 20 grass–clover plots.

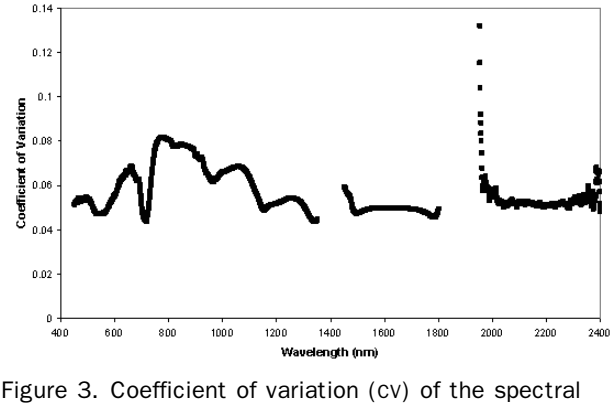

Figure 3. Coefficient of variation (CV) of the spectral measurements for the 20 grass-clover plots.

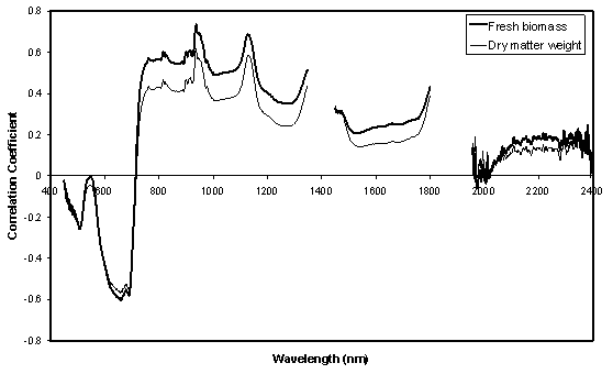

Figure 4. Correlation coefficients relating the canopy reflectance to fresh biomass and dry matter yield, respectively.

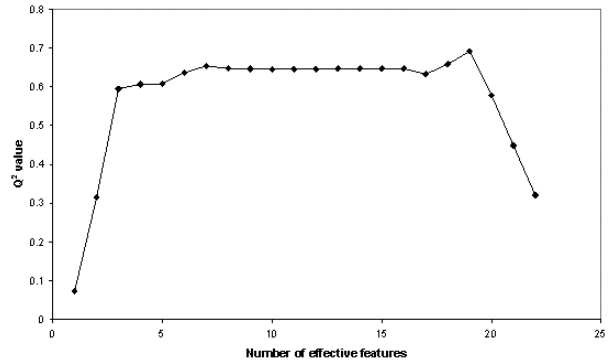

Figure 5. Q^2-results of the band shaving for fresh biomass with a varying number of features (spectral bands) in the prediction model.

spectroradiometer and using a PLS model are shown in Table 2. Based on a randomization t-test only one latent variable had to be used for the PLS model. The PLS model accounted for 55 percent of the variation (Q^2) within the fresh biomass.

Stepwise Forward Selection

Stepwise forward regression applied to the FieldSpec® measurements yielded two spectral bands as optimum result for the fresh biomass. Table 2 provides the results of the stepwise linear regression. Results indicate that, in addition to the 937 nm wavelength band already observed in Figure 4, the red-edge region also provided significant information. For fresh biomass (and also for dry matter weight), this was at the 724 nm band. Noticeable, the latter spectral region on its own does not show a high correlation with these vegetation variables. However, in combination with the 937 nm spectral band it is significant. Other studies also showed the significance of wavelengths at the red-edge region in addition to the NIR region in estimating vegetation biophysical properties (Hansen and Schjoerring, 2003; Schlerf *et al.*, 2005; Thenkabail *et al.*, 2000). Results show a smaller Q^2-value than the Q^2-value obtained with the PLS models, indicating that some information is lost by selecting just these two bands in a simple linear model.

SVM Band Shaving

Subsequently, SVM band shaving was applied as another feature reduction technique. Figure 5 shows the effect of an increasing number of features in terms of Q^2-value for fresh biomass. From this figure we can conclude that the predictive power of a model with three up to 19 features has a Q^2 above 0.6, which can be considered as good for this dataset.

TABLE 2. RESULTS OF THE BIOMASS PREDICTION IN TERMS OF Q^2 AND RMSECV USING PLS (a), STEPWISE FORWARD SELECTION (b) AND SVM BAND SHAVING (c)

	Q^2-value	RMSECV
(a) Partial Least Squares (1 latent variable)	0.55	2.23
(b) Stepwise Regression (bands: 937 nm and 724 nm)	0.48	2.37
(c) Band shaving (3 broad features)	0.59	2.10

For fresh biomass, the three main features were the following wavelength ranges: (a) 668 to 776 nm, (b) 777 to 858 nm, and (c) 859 to 1,006 nm. We see that 724 nm from Table 2 is within the first interval and 937 nm is within the third interval. The same three features as found for fresh biomass were also found as main features for dry matter weight. Table 2 summarizes the observed results. The explained variation in fresh biomass using the three broad bands obtained from shaving is better than the one obtained for one latent variable with PLS. It is noticeable that again the red-edge region is found as a very important feature.

Vegetation Indices

Finally, the range of vegetation indices mentioned before was tested as predictors for fresh biomass. For most traditional indices like SR, NDVI, WDVI, SAVI, GEMI, and EVI, wavelengths of 670 nm and 870 nm were used for the red and NIR bands, respectively. For the EVI also the band at 450 nm in the blue was used. Best result was obtained for the classic SR, whereas results for the NDVI, the MTCI (hyperspectral index) and the NDWI were nearly the same as for the SR. Table 3 provides an overview of the results for the indices. The red-edge indices, particularly those using the contiguous range of wavelengths at the red-edge region (the second

TABLE 3. RESULTS FOR 14 VEGETATION INDICES APPLIED FOR FRESH BIOMASS PREDICTION OF GRASSLAND USING THE WAVELENGTHS GIVEN BY LITERATURE

Vegetation Index	Wavelengths Used (nm)	Q^2-value	RMSECV
SR	670/870	0.47	2.40
NDVI	670/870	0.44	2.47
WDVI	670/870	0.20	2.95
SAVI	670/870	0.22	2.92
GEMI	670/870	0.18	2.98
EVI	450/670/870	0.23	2.90
MTCI	681/709/754	0.45	2.44
RE – Guyot method	670/700/740/780	0.38	2.61
RE – Lagrangian	700 to 740 interval	0.12	3.10
RE – second deriviative	700 to 740 interval	0.11	3.12
WI	900/970	0.33	2.70
NDWI	860/1240	0.43	2.49
NDWI-1640	860/1640	0.37	2.62
NDWI-2130	860/2130	0.22	2.91

TABLE 4. RESULTS FOR FIVE VEGETATION INDICES APPLIED FOR FRESH BIOMASS PREDICTION OF GRASSLAND USING THE 937 NM BAND INSTEAD OF THE NIR 870 NM BAND

Vegetation Index	Wavelengths Used (nm)	Q^2-value	RMSECV
SR	670/937	0.67	1.91
NDVI	670/937	0.65	1.96
WDVI	670/937	0.45	2.45
SAVI	670/937	0.48	2.38
GEMI	670/937	0.45	2.45

TABLE 5. RESULTS FOR FIVE VEGETATION INDICES APPLIED FOR FRESH BIOMASS PREDICTION OF GRASSLAND USING THE 668 TO 776 NM BROAD BAND AS RED BAND AND THE 859 TO 1006 NM BROAD BAND AS NIR BAND RESULTING FROM THE BAND SHAVING PROCEDURE

Vegetation Index	Wavelengths Used (nm)	Q^2-value	RMSECV
SR	668 to 776/859 to 1006	0.44	2.46
NDVI	668 to 776/859 to 1006	0.45	2.45
WDVI	668 to 776/859 to 1006	0.61	2.07
SAVI	668 to 776/859 to 1006	0.49	2.35
GEMI	668 to 776/859 to 1006	0.28	2.80

derivative and the Lagrangian technique), were not the best predictors for the fresh biomass. Actually these were developed for predicting chlorophyll content of the leaves. Also the indices using wavelengths in the SWIR region were not better than the traditional indices using red and NIR wavelengths.

Results could be improved in this study by using the wavelength at 937 nm instead of the one at 870 nm. Of course, this does not apply to the red-edge related indices or to the SWIR related indices. Table 4 now provides the results. Again the SR and NDVI show the best results. The predictive power of the SR and NDVI clearly increased and the Q^2-values are rather high when comparing them with the results of the PLS, the stepwise regression and the band shaving. This may be due to the non-linearity of these indices. Figure 6 illustrates the relationship between the measured fresh biomass and the predicted fresh biomass using the SR.

Another option would be to use the best spectral bands from the band shaving analysis for calculating the vegetation indices. The first and third band obtained covered the two bands obtained with the stepwise linear regression for fresh biomass and dry matter yield. Therefore, we decided to use the 668 to 776 nm band obtained from the band shaving as red band in the SR, NDVI, WDVI, SAVI, and GEMI calculations. The 859 to 1,006 nm band was used as NIR band. The WDVI (Clevers, 1989) appeared to be the best predictor for fresh biomass (Table 5). With these broad spectral bands, it was not possible to derive specific red-edge indices or indices including the SWIR region. Also the EVI could not be calculated since a band in the blue wavelength region was not selected in the shaving. With these two broad bands results were worse than when using the best narrow bands, but still results illustrate the potential of this approach. Results using the first and second band of the band shaving result for fresh biomass yielded slightly worse results.

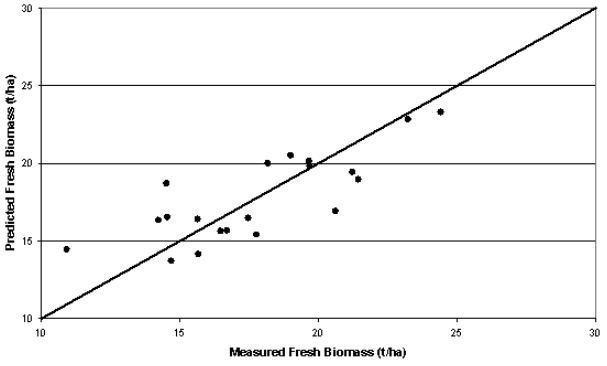

Figure 6. Relationship between measured fresh biomass and predicted fresh biomass using the ratio vegetation index based on the ratio between wavelengths at 937 nm and 670 nm. The bold, straight line depicts the 1:1 line.

Conclusions

In this study, we investigated the optimal spectral bands derived from ASD FieldSpec® spectroradiometer measurements for predicting grassland biomass. Subsequently, the optimal spectral bands were used in a range of vegetation indices as predictors. The potential of a new band shaving technique based on support vector machines (SVM) was tested and compared with a conventional stepwise linear regression and partial least squares (PLS) technique. Results indicate that PLS provides in general a good prediction power. However, it does require the recording of all wavelengths. Stepwise forward regression yielded a narrow band at 937 nm and one at 724 nm as best predictors. Limiting the number of bands to only three broad bands using SVM band shaving yielded good results, which were even slightly better than without band selection. The results indicate that in addition to one band in the NIR region (either 777 to 858 nm or 859 to 1,006 nm) a spectral band in the red-edge region (668 to 776 nm) is important for predicting fresh and dry biomass when using a vegetation index. The broad bands do include the narrow-band results obtained from the stepwise regression. Best result was obtained for the WDVI, but we should note that we tested a lot of vegetation indices and selected the most suitable one from the dataset. This results tentatively in an over-estimation of the prediction power. This is not the case for the (nested) leave-one-out approach applied in PLS and stepwise regression.

Frequent measurements of grass canopy relevant variables during the growing season are a major prerequisite for predicting quality and yield. Future research will focus on the deployment of such technologies allowing for robust estimates with a limited number of spectral bands. Various approaches seem to be feasible having sufficient predictive modeling potential, however little converging of approaches in choosing and proposing appropriate band combinations is seen. This paper proposes to use an advanced algorithm called SVM band shaving to converge to common band sets more quickly.

Perspectives

If only a limited number of spectral bands are needed for estimating grassland variables, a multispectral system may be built with a simplified design that is not only cheaper than a continuous wavelength covering spectrometer, but that can also have a much higher spatial resolution. In addition, when an imaging spectrometer would be used, efficient band elimination procedures could be implemented, accounting

for the potential need to shift spectral band selection under varying phenological phases or background signals. The best spectral bands obtained by the SVM band shaving now would be serious options in a simple multispectral system for predicting the fresh biomass of grasslands. An imaging system (either airborne or spaceborne) could then be used for mapping purposes. For illustrating the potential of such a system, we applied the previous findings to airborne UltraCam™ digital images available for this experiment.

Vexcel's UltraCam$_D$™ digital camera system delivers large format aerial imagery that is radiometrically and geometrically superior to images captured by conventional film cameras at a comparable price (Schiewe, 2005). The UltraCam™ system comprises a panchromatic band and four multispectral bands. On 28 July 2004 the grassland field was recorded using the UltraCam$_D$™ digital camera by Aerodata International Surveys. The flight altitude was about 2,790 m, resulting in a pixel size of 0.25 m for the panchromatic band (450 to 665 nm) and 0.78 m for the multispectral bands (blue: 420 to 475 nm, green: 455 to 580 nm, red: 635 to 675 nm, and NIR: 700 to 805 nm). Aerodata performed a geometric correction for the internal camera geometry and a radiometric correction for vignetting and white balance. The pixel values of the UltraCam™ images were converted to surface reflectance by using reference targets, measured in the same campaign with the FieldSpec® spectroradiometer. Spectral sensitivity curves as provided by Vexcel Corporation were used for calculating the target reflectance for the UltraCam™ spectral bands. An empirical line transformation was used in order to correct for camera calibration and atmospheric effects (Clevers *et al.*, 2005). Average reflectance values per plot were based on about 45 pixels within the plot boundaries.

The spectral bands do not really match the bands obtained from the band shaving method, although the red and NIR band of the UltraCam™ do show correspondence with the first two bands obtained as best from the band shaving. Therefore, the red and NIR bands of the UltraCam™ were also used for calculating the broad-band indices used in this study and the same prediction procedure of fresh biomass was applied as before. Results are given in Table 6. The WDVI offers the best predictive power for these bands. Results are worse than in case of the more optimal bands found before. A linear regression model was then applied to the WDVI obtained from the UltraCam™ images for the entire field, thus providing a prediction map of the fresh biomass. The result is shown in Figure 7. We may improve measurements of field heterogeneity of biophysical variables with remote sensing devices like the UltraCam$_D$™ camera by using more appropriate bands (filters) using band-selection procedures as we have applied in this paper.

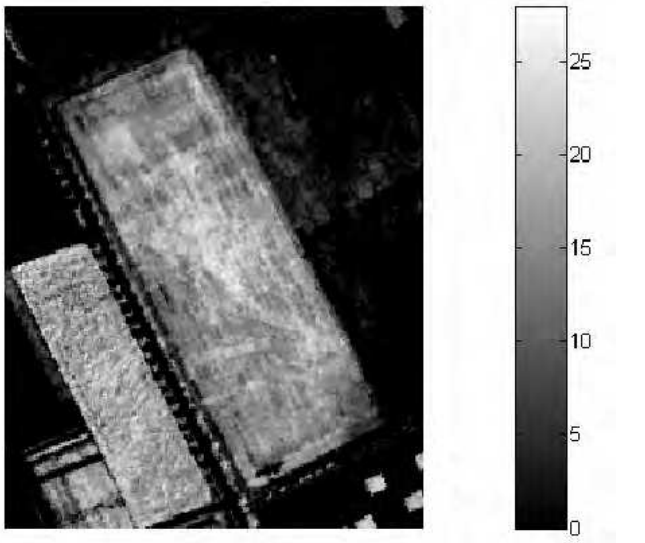

Figure 7. Spatial pattern of the estimated biomass (t/ha) for the grass-clover parcel at the "Droevendaal" experimental farm, Wageningen, based on the spectral information of an UltraCam™ image. The biomass was estimated from the WDVI, obtained from the UltraCam™ image, using the following equation: Biomass (t/ha) = –16.9 + 0.619 ∗ WDVI(%). A color version of this figure is available at the ASPRS website: **www.asprs.org**.

Acknowledgments

We would like to thank Fred Hagman from Aerodata International Surveys for providing the UltraCam™ data and Vexcel Corporation for providing the specifications and spectral sensitivities of the UltraCam™ camera.

TABLE 6. RESULTS OF THE BEST VEGETATION INDICES USING THE RED AND NIR SPECTRAL BANDS OF THE ULTRACAM™ DIGITAL CAMERA SYSTEM.

Vegetation Index	Wavelengths used (nm)	Q^2-value	RMSECV
SR	635 to 675/700 to 805	0.35	2.67
NDVI	635 to 675/700 to 805	0.38	2.61
WDVI	635 to 675/700 to 805	0.40	2.54
SAVI	635 to 675/700 to 805	0.39	2.58
GEMI	635 to 675/700 to 805	0.34	2.69

References

Atzberger, C., 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models, *Remote Sensing of Environment*, 93(1–2):53–67.

Broge, N.H., and E. Leblanc, 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density, *Remote Sensing of Environment*, 76(2):156–172.

Bruzzone, L., and S.B. Serpico, 2000. A technique for feature selection in multiclass problems, *International Journal of Remote Sensing*, 21(3):549–563.

Chen, D.Y., J.F. Huang, and T.J. Jackson, 2005. Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands, *Remote Sensing of Environment*, 98(2–3):225–236.

Clevers, J.G.P.W., 1989. The application of a weighted infrared-red vegetation index for estimating Leaf Area Index by correcting for soil moisture, *Remote Sensing of Environment*, 29(1):25–37.

Clevers, J.G.P.W., and R.E.E. Jongschaap, 2001. Imaging spectrometry for agricultural applications, *Imaging Spectrometry: Basic Principles and Prospective Applications* (F.D. Van der Meer and S.M. De Jong, editors), Kluwer Academic Publishers, Dordrecht, pp. 157–199.

Clevers, J.G.P.W., G.W.A.M. Van der Heijden, and A.G.T. Schut, 2005. A method using different remote sensing techniques for estimating grassland bio-physical variables, *Proceedings of the 4th EARSeL Workshop on Imaging Spectroscopy*, 27–29 April, Warsaw, Poland, pp. 835–849.

Combal, B., F. Baret, M. Weiss, A. Trubuil, D. Mace, A. Pragnere, R. Myneni, Y. Knyazikhin, and L. Wang, 2003. Retrieval of canopy biophysical variables from bidirectional reflectance – Using prior information to solve the ill-posed inverse problem, *Remote Sensing of Environment*, 84(1):1–15.

Curran, P.J., 1989. Remote sensing of foliar chemistry, *Remote Sensing of Environment*, 30(3):271–278.

Dash, J., and P.J. Curran, 2004. The MERIS terrestrial chlorophyll index, *International Journal of Remote Sensing*, 25(23):5403–5413.

Daughtry, C.S.T., C.L. Walthall, M.S. Kim, E. Brown de Colstoun, and J.E. McMurtrey III, 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance, *Remote Sensing of Environment*, 74(2):229–239.

Dawson, T.P., and P.J. Curran, 1998. A new technique for interpolating the reflectance red edge position, *International Journal of Remote Sensing*, 19(11):2133–2139.

Demetriadesshah, T.H., M.D. Steven, and J.A. Clark, 1990. High-resolution derivative spectra in remote sensing, *Remote Sensing of Environment*, 33(1):55–64.

Elvidge, C.D., and Z.K. Chen, 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indexes, *Remote Sensing of Environment*, 54(1):38–48.

Gao, B.C., 1996. NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment*, 58(3):257–266.

Gong, P., R.L. Pu, G.S. Biging, and M.R. Larrieu, 2003. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1355–1362.

Guyon, I., J. Weston, S. Barnhill, and V. Vapnik, 2002. Gene selection for cancer classification using support vector machines, *Machine Learning*, 46(1–3):389–422.

Guyot, G., and F. Baret, 1988. Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux, *Proceedings 4th International Colloquium 'Spectral Signatures of Objects in Remote Sensing'*, 18–22 January 1988, Aussois, France, pp. 279–286.

Haboudane, D., J.R. Miller, E. Pattey, P.J. Zarco-Tejada, and I.B. Strachan, 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture, *Remote Sensing of Environment*, 90(3):337–352.

Haboudane, D., J.R. Miller, N. Tremblay, P.J. Zarco-Tejada, and L. Dextraze, 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture, *Remote Sensing of Environment*, 81(2–3):416–426.

Hansen, P.M., and J.K. Schjoerring, 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression, *Remote Sensing of Environment*, 86(4):542–553.

Hastie, T., R. Tibshirani, and J. Friedman, 2001. *The Elements of Statistical Learning*, Springer, New York, 522 p.

Huang, Z., B.J. Turner, S.J. Dury, I.R. Wallis, and W.J. Foley, 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis, *Remote Sensing of Environment*, 93(1–2):18–29.

Huete, A., K. Didan, T. Miura, E.P. Rodriguez, X. Gao, and L.G. Ferreira, 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sensing of Environment*, 83(1–2):195–213.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, 25(3):295–309.

Hunt, E.R., 1991. Airborne Remote-Sensing of Canopy Water Thickness Scaled from Leaf Spectrometer Data, *International Journal of Remote Sensing*, 12(3):643–649.

Jacquemoud, S., C. Bacour, H. Poilve, and J.P. Frangi, 2000. Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode, *Remote Sensing of Environment*, 74(3):471–481.

Kooistra, L., E.A.L. Salas, J. Clevers, R. Wehrens, R. Leuven, P.H. Nienhuis, and L.M.C. Buydens, 2004. Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains, *Environmental Pollution*, 127(2):281–290.

Moran, M.S., Y. Inoue, and E.M. Barnes, 1997. Opportunities and limitations for image-based remote sensing in precision crop management, *Remote Sensing of Environment*, 61(3):319–346.

Penuelas, J., J. Pinol, R. Ogaya, and I. Filella, 1997. Estimation of plant water concentration by the reflectance water index WI (R900/R970), *International Journal of Remote Sensing*, 18(13):2869–2875.

Pinty, B., and M.M. Verstraete, 1992. Gemi – A nonlinear index to monitor global vegetation from satellites, *Vegetatio*, 101(1):15–20.

Rouse, J.W., R.H. Haas, J.A. Schell, D.W. Deering, and J.C. Harlan, 1974. *Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation*, NASA/GSFC, Type III, Final Report, Greenbelt, Maryland, 371 p.

Schiewe, J., 2005. Status and future perspectives of the application potential of digital airborne sensor systems, *International Journal of Applied Earth Observation*, 6(3–4):215–228.

Schlerf, M., C. Atzberger, and J. Hill, 2005. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data, *Remote Sensing of Environment*, 95(2):177–194.

Serrano, L., J. Penuelas, and S.L. Ustin, 2002. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals, *Remote Sensing of Environment*, 81(2–3):355–364.

Thenkabail, P.S., R.B. Smith, and E. De Pauw, 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics, *Remote Sensing of Environment*, 71(2):158–182.

Thenkabail, P.S., R.B. Smith, and E. De Pauw, 2002. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization, *Photogrammetric Engineering & Remote Sensing*, 68(6):607–621.

Thissen, U., M. Pepers, B. Ustun, W.J. Melssen, and L.M.C. Buydens, 2004. Comparing support vector machines to PLS for spectral regression applications, *Chemometrics and Intelligent Laboratory Systems*,73(2):169–179.

Van der Voet, H., 1994. Comparing the predictive accuracy of models using a simple randomization test, *Chemometrics and Intelligent Laboratory Systems*, 25:313–323.

Vapnik, V., 1998. *Statistical Learning Theory*. Wiley, New York, 768 p.

Verzakov, S., P. Paclík, and R.P.W. Duin, 2004. Feature shaving for spectroscopic data, *Lecture Notes in Computer Science*, 3138:1026–1033.